DOCUMENT RESUME

ED 386 065 FL 023 230

AUTHOR Berry, Vivien
TITLE The Assessment of Spoken Language under Varying Interactional Conditions.
PUB DATE 94
NOTE 19p.; In: Bird, Norman, Ed., And Others. Language and Learning. Papers presented at the Annual International Language in Education Conference (Hong Kong, 1993); see FL 023 205.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Students; *English for Academic Purposes; *Extraversion Introversion; Foreign Countries; Group Dynamics; Higher Education; *Interpersonal Communication; Interviews; Language Research; Language Tests; *Oral Language; *Personality Traits; Testing; *Verbal Tests

ABSTRACT

Two studies of individuals' oral second language performance in interaction with extraverts and introverts are reported here. The first, described briefly, investigated the effects of homogeneous (extravert/extravert or introvert/introvert) vs. heterogeneous pairings on oral performance in interviews. Subjects were 36 women students in a Japanese college. As predicted, introverts performed best in homogeneous pairs and somewhat less well in heterogeneous pairs, while extraverts performed best in homogeneous pairs; neither group performed as well in individual interviews. The second study investigated the hypothesis that individual learners of a second language would perform differently on a group oral test depending on the degree of extraversion of the individual in relation to economics students in an English enhancement course. Students had been assigned randomly to seminar groups, which were then rated for extraversion/introversion. Results indicate that the degree of extraversion in the group produced no significant differences in the scores of extraverts, while those of introverts were considerably affected. These apparently contradictory findings are discussed, particularly as they relate to a trend toward paired and group teaching. Contains 48 references. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 386 065

THE ASSESSMENT OF SPOKEN LANGUAGE UNDER VARYING INTERACTIONAL CONDITIONS

VIVIEN BERRY

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Lo Clark

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

FL023233

BEST COPY AVAILABLE

THE ASSESSMENT OF SPOKEN LANGUAGE UNDER VARYING INTERACTIONAL CONDITIONS

Vivien Berry

Abstract

Paired interviews and group discussions are becoming increasingly popular as methods of assessing spoken language. Yet recent research has shown that extreme extraverts and introverts differ in how well they perform on oral test interviews depending on whether personality types are homogeneously or heterogeneously paired. There is also experimental evidence that extraverts and introverts perform differently when tested in groups. This paper will report on a study in which approximately 100 undergraduate students were tested on their ability to take part in an academic seminar. Each student was rated by two experienced raters on a nine point scale. Ratings of speaking performance of both extremes on the extraversion scale (as measured by the EPQ) are compared to the degree of homogeneity of personality type present in each group. Initial results indicate that differences can be observed in the performances of extraverts and introverts under varying interactional conditions. The findings from this research clearly demonstrate the importance of deriving hypotheses from the psychological literature when investigating the effect of personality variables on performance. The paper concludes with a discussion of the feasibility of oral testing in groups and of the stability of results obtained.

Introduction

Of all the skills involved in learning a language, spoken language is the most difficult to assess. It is the most labour intensive and the most time consuming. Speaking is probably also the most difficult skill to score accurately and consequently scores obtained on oral language tests may not necessarily be reliable. Many factors can affect language test scores, among them differences in learners' cultural backgrounds (Chen and Henning 1985, Zeidner 1986, 1987), prior knowledge (Alderson and Urquhart 1985, Hansen and Jenson 1993), gender and academic status (Porter 1990, Cushing 1993, Zammit 1993), the extent of interviewer accommodation (Ross 1992) and different rater characteristics (Elder 1993, Pollitt and Murray 1993).

In the United States, for at least the past decade, the focus of nearly all research related to the assessment of spoken language has been the oral interview, in particular the ILR/ACTFL oral interview and its associated guidelines for the assessment of oral proficiency. This research has been primarily statistical in nature and the over-riding concern has been to provide evidence of the validity of the interview as an instrument to measure spoken language. However, the test format has been criticised for (among other reasons) not accurately reflecting the realistic

features of natural communication (Bachman and Savignon 1986, Bachman 1988), or conversation (van Lier 1989).

Recognition of the shortcomings of some of the features of the interview-as-test has led official examinations organisations such as the University of Cambridge Local Examinations Syndicate (UCLES) and the Royal Society of Arts (RSA) - now amalgamated - and many university second language placement programmes to experiment with variations in both oral interview formats and oral test formats in general. One of the major innovations has been the introduction of paired or group interactions[1] between testees, rather than restricting language interaction to the traditional dyad of interviewer-interviewee. If care is taken in the allocation of learners to pairs or groups, this learner-centred approach to testing has the advantage of reducing, if not altogether removing, some of the tensions associated with the traditional dyad. For example, non-linguistic factors such as ethnicity, gender and social status, all of which have been mentioned by Brindley (1991) as potentially affecting judgements of proficiency, can be controlled for.

Without according it any special status in the hierarchy, Brindley (1991:156) also includes personality as one other non-linguistic factor amongst those he sees as important. Unfortunately, personality is a variable which cannot be controlled for on a simple observational basis. It is maintained in this paper that unless an appropriately validated instrument is used to assess personality, and allocation of learners to pairs or groups is made on a principled basis, taking into account the findings of theoretically sound empirical research, then it is misleading, to say the least, to suggest that personality has been controlled for.

Unfortunately, theoretically sound research findings into the effect of personality characteristics on second language task performance are very hard to find. The problem seems to be that specific hypotheses, derived from the specialist psychological literature, have seldom been formulated. The reason for this is that such hypotheses are not easily identified and they cannot be deduced from the second language literature. Major reviews of the role played by personality variables in second-language learning (Ellis 1986, Skehan 1989) have reached extremely pessimistic conclusions, particularly with regard to the implications of extraversion, as a variable. They point out that many studies have failed to produce any significant findings, citing, for example, Naiman et al. (1978), who failed to find a significant effect for extraversion in characterising the good language learner. It can be argued, however, that the problem lies not so much with the lack of significance of the results obtained but rather that these pessimistic conclusions have been reached through reviewing research which tested hypotheses that are neither logically derived from personality theory, nor predicted from relevant experimental evidence.

Another study, described by Brown as "... the most comprehensive study to date on extroversion." (Brown 1987:110) is that of Busch (1982), who also failed to find support for her somewhat extraordinary hypothesis (hunch?) that "extraverts are more proficient in English." (Busch 1982:109). More recently, Porter conducted

research into affective reactions of learners based on a "rough categorisation of their personalities into 'more outgoing' or 'more reserved'..." (Porter 1991:97). Totally unsurprisingly, he also found that personality type did not seem to have any significant effect. It is findings from theoretically unsound research designs such as those of Busch and Porter, who have adapted psychological constructs merely to test things "which intuitively strike them as important" (Ellis 1986:120), that has led some researchers to reject personality as a significant factor in second language acquisition. However, summarising a comprehensive review of second-language personality studies, Griffiths concludes, "... the fact that researchers have not found relationships cannot be fairly used (as it has been) to dismiss personality variables from the L2 research agenda; nor can highly validated psychometric instruments be held accountable for the failure." (Griffiths 1991:68).

Personality measurement

The major personality dimensions are represented in almost all large scale studies and nearly all theoretical formulations. They are represented by continua, the extremes of which can be described through idealised types:

> Extraverts are sociable, like parties, have many friends and need excitement; they are sensation seekers and risk-takers, like practical jokes and are lively and active. Conversely introverts are quiet, prefer reading to meeting people, have few but close friends and usually avoid excitement. (Eysenck and Chan 1982:154)

A number of instruments have been developed which attempt to measure the major dimensions of personality, amongst them Cattell's 16PF (Cattell et al. 1970) and the Minnesota Multiphasic Personality Inventory (MMPI, Hathaway and McKinley, n.d.). Neither of these has been validated[3] for use in any non-western country and one of the major difficulties is that concepts like Introversion-Extraversion (common to all of them) which have "... an agreed meaning in one culture may not have the same, or indeed any, meaning in another culture." (Iwawaki et al. 1980:195).

When a test is used in a culture other than the one it was originally developed for, evidence of the test's reliability and validity in the new setting is required. Research has shown that reanalysis of culturally transposed tests is needed at the item level in order to identify items that function differentially for the two groups (see Ellis et al. 1993 for a detailed discussion of cross-cultural validation studies using IRT analysis). The importance of cross-cultural validation studies has been pointed out by S.B.G. Eysenck who argues that "... it is imperative that all items be tested for appropriateness before inclusion in any foreign scoring key", whilst warning of the dangers of "spurious results" if this is not done (Eysenck 1983:381)

The psychometric instruments used to assess degrees of extraversion in the studies reported here were the 86-item Japanese version of the Eysenck Personality

Questionnaire, (Iwawaki et al. 1980) and the 90 item Hong Kong EPQ (Eysenck and Chan 1982), both validated for use in the respective countries. What this means in practice is that both instruments had been subjected to translation into Japanese or Cantonese as appropriate, followed by back-translation to iron out obvious translation errors. Once translation errors had been identified and corrected, a content analysis was performed by means of inter-item correlations followed by principal component factor analysis with varimax rotation to simple structure and a final promax rotation to oblique simple structure using only the first four factors for rotation. Only items which loaded solely on one factor were included in the foreign scoring keys thus producing tests which possessed the property of measurement equivalence where "... individuals with equal standing on the trait measured by the test but sampled from different sub-populations have equal expected observed test scores" (Drasgow 1989:19).

It is appropriate to note that methods of validation of the EPQ have been criticised, most notably for the methods used to derive indices of factor comparison (Bijnen et al. 1986). Since the metric assumptions inherent in factor analysis may not be met in real data, the results, particularly of hierarchical factor analysis, may be prone to error. Non-metric multidimensional scaling,[4] which requires only ordinal assumptions of the data, offers a more robust model for multivariate analysis and may be more appropriate for analysis of the item structure of psychological tests like the EPQ. For example, two people might obtain exactly the same scores on the extraversion scale, but have achieved them by giving positive responses to different stimuli. In other words, "extraversion" is composed of more than one underlying dimension. The objective of multidimensional scaling is to determine the number of dimensions differentiating the stimuli (in this case the items in the EPQ). Individual stimuli are represented by points in geometric space; the more similar the stimuli, the closer the points. Smallest space analysis[5] of the item structure of the EPQ shows that the extraversion items form a "tight cluster" (Hammond 1987:545), thus providing further psychometric validation of the E-scale. A full discussion of the criticisms, defences and validation procedures of the EPQ is beyond the scope of this paper. However recent studies (Hanin et al. 1990) accept these criticisms and now state their results not in terms derived solely from factor analysis, but also from multidimensional scaling using smallest space analysis (Lingoes 1973).

It is clear that within the psychological community the EPQ has provoked both much criticism and a substantial body of supportive research. With the exception of the best-known IQ tests, it is probably one of the most extensively researched measurement instruments in existence. Even if philosophical doubts exist concerning the trait structure of the EPQ and the dimensions of personality it is measuring, the numerous, methodologically sound validation procedures, carried out in over forty countries over as many years, support the existence of a stable notion of extraversion which is relatively invariant, replicable and, more importantly, subject to falsification.

The problem of establishing the construct validity of the EPQ (does it measure what it is intended to measure) is, of course, circular in that there is no external criterion against which the test can be evaluated since the existence of such a criterion would make the test itself unnecessary! The only way that the validity of the EPQ can be established other than statistically, is by deriving hypotheses logically predicted from the theory, testing them and determining if they fit the predictions. A review of the experimental research reported in the psychological literature reveals several studies where it is not only possible to draw meaningful hypotheses, but also to relate them specifically to the methods of L2 testing currently under consideration. The remainder of this paper will present evidence from two such studies, the results of which show that significant differences can be observed in the responses of introverts and extraverts under varying interactional conditions.

Study 1. Paired interactions on an interview test

In the first study, extensively reported elsewhere (Berry 1993), the present researcher investigated the hypothesis, derived from Leith (1974) and further supported by the findings of Hall et al. (1988), that there would be significant differences in performance of both introverts and extraverts on an oral interview test, dependent on method of pairing. Specifically, it was predicted (again from Leith 1974) that introverts would perform best if interviewed in homogeneous pairs, next best if interviewed as individuals and worst if interviewed in heterogeneous pairs. Extraverts, on the other hand, would again perform best in homogeneous pairs but would do next best in heterogeneous pairs and worst as individuals (see Table 1). Unlike second language personality studies which generally obtain findings based on global correlational measures, psychological research in this area usually compares selected groups of extreme introverts and extreme extraverts (Cook 1993:91), "extreme" meaning plus or minus one standard deviation or more from the mean.

36 second year female students from a Japanese junior college (18 each of extreme E and I) took part in the study. No significant differences were found in their levels of general language proficiency as measured by an Institutional TOEFL. Students were randomly assigned to each of the three possible personality pairings and 24 interviews were conducted as follows (Table 2): 6 individual interviews of both I and E (12), 3 homogeneously paired interviews of both I and E (6) and 6 heterogeneously paired interviews (6). This allowed for a total of six sets of scores to be analyzed in each of the six possible categories (I-individual, E-individual, I+I, E+E, I+E, E+I). Interviews of the different categories were also conducted in random order.

Table 1

Achievements of Students Learning in Homogeneous or Heterogeneous Personality Pairs or as Individuals (from Leith 1974)

Personality	Methods		
	Homogeneous pairs	Heterogeneous pairs	Individuals
Introverts	32.2	27.3	30.0
Extroverts	30.6	27.7	25.4
Significance of Differences	n.s.	n.s.	p<.01

Homogeneous vs heterogeneous pairs: p<.01
Homogenous pairs vs individuals: p<.025

Table 2

Methods of Pairing for Interviews

Personality type	Number of interviews	Number of students
Individual I	6	6
Individual E	6	6
I+I	3	6
E+E	3	6
I+E	6	12 (6I+6E)
Total	24	36

The test itself consisted of a four part interview designed to approximate the level and format of the Cambridge Preliminary English Test (PET). Means of overall averages were calculated for each of the categories of interviews. Analysis of means yielded the following results:

Table 3

Comparison of Scores on Oral Interview Tests

Personality	Methods		
	Individuals	Homogeneous pairs	Heterogeneous pairs
Introverts	61.15	69.80	68.33
Extroverts	56.04	80.21	71.35
Significance of Differences	n.s.	$p<.05$	n.s.

Introverts: Homogeneous vs heterogeneous pairs: n.s.
Individuals vs both pairs: $p<.05$
Extraverts: All results $p<.05$

The results provide partial support for the original hypotheses. Extraverts performed exactly as predicted, showing dramatic increases over individual performance when interviewed in pairs and performing best of all in homogeneous pairs. Scores for introverts are highest in homogeneous pairs although these are not significantly different from those of heterogeneous pairs. However, against expectations, scores on individual interviews are significantly lower than on either of the pairings, suggesting that variables other than extraversion are having an effect. In fact, as both introverts and extraverts do least well in an individual interview, it may be that culturally stereotypic views of the interviewer-interviewee relationship are disturbed by, for example, having to interact in a role-play situation. Nevertheless, given the small sample size, the results are interesting and certainly indicate that further research in this area is necessary before testing in pairs is adopted wholesale.

Study 2: Participation in a group oral test.

This study investigated the hypothesis that individual learners would perform differently on a group oral test depending on the degree of extraversion of an individual in relation to the amount of extraversion present within the group.

The theoretical background for this study can be found in the work of Jennifer George (1990) who explored personality, affect and behaviour as group level phenomena in relation to absenteeism at work. She found considerable support for her hypothesis that characteristic levels of the personality traits PA (positive affect)

and NA (negative affect), within work groups would be related to the positive and negative affective tones of the groups respectively. PA and NA, measured by using the appropriate scales of the Multidimensional Personality Questionnaire (MPQ, Tellegen 1982, cited in George 1990), have been shown to be related to the extraversion scale of other personality measures (George 1990:109). Characteristic levels of NA and PA within groups were determined by averaging group-member scores. Group affective tone was measured by aggregating the individual measures obtained on the Job Affect Scale (JAS, Brief et al. 1988, cited in George 1990).

Background and description of group oral test

Unlike the previous study which attempted to control as many variables as possible in an experimental research design, the setting for the current study was firmly grounded in the real world, with all the attendant constraints thus implied. The population sample was drawn from first year Economics students entering the University of Hong Kong. All first year Economics students are required by their department to take a twenty week English Enhancement course taught in the English Centre. Before the course starts, they are given an oral test, since until this year there has been no oral component in the Use of English Examination. The aim of the oral test is to provide an opportunity for students to interact with their peers in an authentic university setting , thus providing samples of language, the assessment of which provides meaningful information for both students and teachers.

The test format is designed to replicate, as closely as possible, the setting of a small academic seminar[6]. Students are assigned to groups, generally with five to a group (on the basis of their English Centre registration number which has been assigned alphabetically). They are given a short text to read, allowed five minutes to take notes on it, then asked to discuss it seriously on two levels: 1) in relation to the research and information given, and 2) by relating the research findings to their own experience and the situation in Hong Kong. Each group is assessed by two teachers - one who acts as tutor by starting off the discussion (subsequently tutors intervene only if all communication has broken down) and one who acts as an observer and takes no direct part in the proceedings. Both teachers individually assess each student using a nine point letter scale for each of: relevance / participation and articulation. At the end of each 'seminar' session, teachers discuss the grades given and agree on one grade for each category for each student.

After taking part in the oral seminar assessment exercise, each student was asked to complete a personality questionnaire. The instrument used to assess degrees of extraversion was the 90-item Hong Kong Chinese version of the EPQ (Eysenck Personality Questionnaire, Eysenck and Eysenck 1975) which emerged from the cross-cultural validation studies carried out by Eysenck and Chan in 1982 and which was kindly supplied for this study by Dr. Chan. The means and standard deviations obtained on the extraversion scale for the university sample were

generally similar to those obtained by Eysenck and Chan 1982 although it will be noted that the mean for extraversion is slightly lower for the current sample.

Table 4

Comparison of EPQ Scores With Eysenck and Chan (1982)

Study	Sex	mean	s.d.	n
Eysenck and Chan 1982	(male)	12.17	4.43	270
	(female)	11.24	4.44	462
H.K.U. students 1993	(male)	11.25	4.23	38
	(female)	10.97	4.66	64

On the basis of their responses on the EPQ, students were classified as either extreme extravert, extreme introvert or ambivert. The number of extremes is interesting since the percentage in Japan was approximately 35% whereas in Hong Kong exactly 50% were classified as extremes. Obviously the higher the percentage of the population classed as extremes, the more important research is into how individual differences in personality affect performance.

Table 5

Distribution of Personality Types on Extraversion Scale of EPQ

	Extravert	Introvert	Ambivert	Total
n	26	25	51	102
mean E score	17.65	5.64	11.22	
s.d.	1.09	1.41	1.9	

As mentioned previously, students were assigned to their seminar groups quasi-randomly on the basis of their university registration numbers. The degree of extraversion present in each group was determined by averaging the E scores of each member of the group. One group consisted entirely of ambiverts (within one standard deviation of the mean in either direction) which left 20 groups for analysis. Categories of groups were then established as follows:

Table 6

Categorisation of Groups by Mean Level of Extraversion.

Category	Mean extraversion in group
1	≥ 13
2	$12 < 13$
3	$11 < 12$
4	$10 < 11$
5	$9 < 10$
6	< 9

Each category was then individually inspected to determine placement of individual extraverts and introverts within them. To control for possible differences in general language proficiency H.K.E.A. Use of English results were compared. Means were calculated for each category and analysis of means revealed no significant differences.

The results reported in Table 7 give some support to the original hypothesis that there would be observable differences in the performance of extraverts and introverts depending on the degree of extraversion present in the group. They do not, of course provide overwhelming evidence for the role of extraversion in group interaction.

However, there is a trend, supported by the finding of significant differences in the means between categories 1 and 6 that Introverts are affected by the degree of extraversion present in a group, whereas extraverts are not. It would seem that when placed in a group with a relatively high degree of extraversion, introverts respond positively to the group dynamics and therefore are rated more highly, at least for relevance/participation (no significant differences were observed between any groups for articulation) When placed in a group with a lower degree of

introversion, individual introverts remain quiet and are therefore rated less highly.

Table 7

Mean Scores of Extreme E and Extreme I by Category (for Relevance / Participation)

Category	Mean Group Extraver -sion	Mean Oral score (I)	n (25)	Mean oral score (E)	n (26)
1	·13≥	6.5	3	4.5	5
2	12 <13	5.6	4	5.6	6
3	11 <12	5.4	7	5.4	8
4	10 <11	5.7	4	5	7
5	9 <10	4.3	4		
6	<9	3.3	3		

Significance of differences: Extravert vs Introvert = n.s.
Extravert Groups 1-6 = n.s.
Introvert Groups 1 and 6 $p < 0.05$

Implications

The results of the studies discussed cannot be considered to provide conclusive evidence of bias either in favour of or against any particular personality type. Until they can be replicated on a much larger scale, they can only suggest potential problems of interpretation of scores. It is however interesting to note the apparently contradictory findings of the two studies with respect to the effects of extraversion. In the first study, extraverts did much better when placed in homogeneous pairs than in heterogeneous pairs whereas there were no significant differences for introverts. However, in the second study, the degree of extraversion present in a group produced no significant differences in the scores of extraverts whereas introverts were considerably affected. One possible explanation is that method has a tremendous effect on how extraverts and introverts perform. For example, there is a considerable body of research evidence which indicates that introverts are favoured by a well-structured, highly prompted learning situation (the PET is an extremely prescriptive, structured test) while extraverts are better off when

presented with a high degree of uncertainty and ambiguity, such as the seminar situation (e.g. Shadbolt 1978, Riding and Parker 1979). It may be that the method effect is dominant and differences are only observed when either extreme is placed in their least favoured situation.

Given the direction towards pair and group testing by influential testing boards, this area of research could well prove to be of major importance in the very near future. Small-stakes tests are, of course, not important. Placing a student in the wrong level of class is instantly rectifiable. But what of the introverted students who turn up for the new H.K.E.A.Use of English oral exam. and find themselves placed in groups with several other introverts? What if those small differences in scores are norm-referenced so that one of them receives, for example, an E9 instead of a D8? The University of Hong Kong has an admissions policy which puts the cut-off entry point at Grade D8, so any student in the situation outlined above would be refused admission. That is when the stakes get very high indeed.

One final comment is perhaps appropriate. Even if personality characteristics are innate (and this is not altogether uncontentious), it may be that extraverts and introverts use different strategies to cope with the identical situations they both have to face in every day life. This is a very promising area for research since it adds a human dimension to the psychometric validity issues. There is at least a possibility that if differences in strategy use can be established, the problem of potential test bias can to a certain extent be overcome by appropriate learner training.

Notes

1. This is, of course, not an 'innovation' in Israel where the idea of group oral examinations dates back to at least 1980 (Reves 1980, Reves 1982, Shohamy, Reves and Bejarano 1986).

2. This paper will maintain the spelling of extraversion generally used in the psychological literature. When quoting other sources directly, the spelling used by each particular author will be adopted

3. It is important to distinguish between 'translation' by which items from a scale are translated into another language and 'validation' where items are subjected to statistical analysis (usually factor or smallest space analysis) before being included in a foreign version of a test.

4. There are two stages to multidimensional scaling. The first step is to determine the number of dimensions underlying whatever phenomenon is under investigation. The second step is to obtain scale values for the stimuli on a selected set of

dimensions. For a fuller description of the principles of multidimensional scaling and of the procedures involved, see Nunnally 1978, Chapter 2.

5. Described by Hammond (1987:544) as "One of the most elegant multidimensional scaling algorithms...", smallest space analysis was originally proposed by Louis Guttman (1968).

6. For a full description of the rationale and development of this test at the University of Hong Kong, see Morrison and Lee 1985.

Acknowledgements

The research reported in the second study was supported by research grant #335/091/0001 (C.R.C.G.). I would like to thank the students and teachers in Japan and Hong Kong who participated in these studies. In particular I would like to acknowledge the contribution of Dr. Roger Griffiths to the Japanese study. Special thanks are also due to Dr. Desmond Allison for his insightful and thought-provoking comments on earlier versions of this paper.

References

Alderson, J.C. & Urquhart, A. (1985). The effect of students' academic discipline on their performance on ESP reading tests. *Language Testing* 2(2):192-204.

Bachman, L.F. (1988). Problems in examining the validity of the ACTFL oral interview. *Studies in Second Language Acquisition* 10(2):149-164.

Bachman, L.F. & Savignon, S.J. (1986). The evaluation of communicative language proficiency: a critique of the ACTFL oral interview. *Modern Language Journal* 70(4):380-390.

Bachman, L.F., Lynch, B. & Mason, M. (1993). Investigating variability in tasks and rater judgements in a performance test of foreign language ability. Paper presented at the 15th Language Testing Research Colloquium, Arnhem & Cambridge. August 1993.

Berry, V. (1993). Personality characteristics as a potential source of language test bias. In A. Huhta, K. Sajavaara, & S. Takala. (eds.) *Language Testing: New Openings*, (pp. 115-124). Jyvaskyla, Finland: Institute for Educational Research.

Bijnen, E.J., Van der Net, T.Z. & Poortinga, Y.H. (1986). On cross-cultural comparative studies with the Eysenck Personality Questionnaire. *Journal of Cross-Cultural Psychology* 17(1):3-16.

Brindley, G. (1991). Defining language ability: the criteria for criteria. In S.Anivan (ed.) *Current Developments in Language Testing*. Anthology Series 25. (pp. 139-164). Singapore:SEAMEO.RELC.

Brown, H.D. (1987). *Principles of Learning and Teaching*. Englewood Cliffs, NJ.: Prentice-Hall.

Busch, D. (1982). Introversion-extraversion and the EFL proficiency of Japanese students. *Language Learning* 32(1):109-132.

Cattell, R.B., Eber, H.W. & Tatsouka, M.M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire (16PF)*. Champaign, Illinois: Institute for Personality and Ability Testing.

Chen, Z. & Henning, G. (1985). Linguistic and cultural bias in language proficiency tests. *Language Testing* 2(2):155-63.

Cook, M. (1993). *Levels of Personality*. (2nd edition). London: Cassell.

Cushing, S.T. (1993). L2 Proficiency, Academic Status, and Lecture Note Content. Paper presented at the TESOL Conference, Atlanta, GA. April 1993.

Drasgow, F. (1987). Study of the measurement bias of two standardized psychological tests. *Journal of Applied Psychology* 72(1):19-29.

Elder, C. (1993). Are raters' judgements of language teacher effectiveness wholly language-based? Paper presented at the 15th Language Testing Research Colloquium, Arnhem & Cambridge. August 1993.

Ellis, B.B., Becker, P. & Kimmel, H.D. (1993). An item response theory evaluation of an English version of the Trier Personality Inventory (TPI). *Journal of Cross-cultural psychology* 24(2):133-148.

Ellis, R. (1986). *Understanding Second Language Acquisition*. Oxford: Oxford University Press.

Eysenck, S.B.G. (1983). One approach to cross-cultural studies of personality. *Australian Journal of Psychology* 35(3):381-91.

Eysenck, S.B.G. & Chan, J. (1982). A comparative study of personality in adults and children: Hong Kong vs England. *Personality and Individual Differences* 3(3):153-160.

Eysenck, H.J. & Eysenck, S.B.G. (1975). *Manual of the Eysenck Personality Questionnaire*. London: Hodder and Stoughton.

George, J.M. (1990). Personality, affect and behaviour in groups. *Journal of Applied Psychology* 75(2):107-116.

Griffiths, R.T. (1991). Language Learning and the Japanese Personality: Hypotheses from the Psychological Literature. In T. Hayes & K. Yoshioka (eds.), *Proceedings of the Conference on Second Language Research in Japan*. (pp. 61-121). Tokyo:The International University of Japan.

Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrica* 33:469-506.

Hall, R.H., Rocklin, T.R., Dansereau, D.F., Skaggs, L.P., O'Donnell, A.M., Lambiotte, J.G. & Young, M.D. (1988). The role of individual differences in the cooperative learning of technical material. *Journal of Educational Psychology* 80(2):172-178.

Hammond, S.M. (1987). The item structure of the Eysenck Personality Questionnaire across method and culture. *Personality and Individual Differences* 8(4):541-549.

Hanin, Y., Eysenck, S.B.G., Eysenck, H.J. & Barrett, P. (1991). A cross-cultural study of personality: Russia and England. *Personality and Individual Differences* 12(3):265-71.

Hansen, C. & Jenson, C. (1993). The effect of prior knowledge on EAP listening test performance.. Paper presented at the 15th Language Testing Research Colloquium, Arnhem and Cambridge. August 1993.

Hathaway, S.R. & McKinley, C. (n.d.). *The Minnesota Multiphasic Personality Inventory (MMPI): New Group Form (Form R)*. Minnesota: Minnesota University Press.

Iwawaki, S., Eysenck, S.B.G. & Eysenck, H.J. (1980). Japanese and English personality structure: a cross-cultural study. *Psychologia* 23(4):195-205.

Leith, G.O.(1974). Individual differences in learning: Interactions of personality and teaching methods. Personality and Academic Progress proceedings.(pp. 14-25). London:Association of Educational Psychologists.

Lingoes, J.C. (1973). *The Guttman-Lingoes nonmetric program series*. Ann Arbor: Mathesis.

Morrison, D.M. & Lee, N. (1985). Simulating an academic tutorial: a test validation study. In Y.P.Lee, A.C.Y.Y. Fok, R.Lord & G.Low (eds.) *New Directions in Language Testing*, (pp. 85-92). Oxford: Pergamon Press.

Naiman, N., Frohlich, M., Stern, H.H. & Todesco, A. (1978). *The Good Language Learner. Research in Education Series #7.* Toronto: Ontario Institute for Studies in Education.

Nunnally, J.C. (1978). *Psychometric theory*. (2nd edition) New York: McGraw Hill.

Pollitt, A. & Murray, N.L. (1993). What raters really pay attention to. Paper presented at the 15th Language Testing Research Colloquium, Arnhem and Cambridge. August 1993.

Porter, D. (1991). Affective factors in the assessment of oral interaction: gender and status. In S.Anivan (ed) *Current Developments in Language Testing*. Anthology Series 25. (pp. 92-102). Singapore: SEAMEO.RELC.

Reves, T. 1980. The group-oral Bagruth examination: an experiment. *English Teachers Journal* 24. (quoted in Reves 1982).

Reves, T. 1982. The group-oral examination: a field experiment. *World Language English* 1(4):259-262.

Riding, R.J. & Parker, J.E. (1979). The effect of extraversion, detail importance and interference on the recall of prose by eleven-year old children. *Educational studies*, 5(1):15-22.

Ross, S. (1992). Accomodative questions in oral proficiency interviews. *Language Testing* 9(2):173-186.

Shadbolt, D.R. (1978). Interactive relationships between measured personality and teaching strategy variables. *British Journal of Educational Psychology*, 48(2):227-231.

Shohamy, E., Reves, T. & Bejarano, Y. (1986). Introducing a new comprehensive test of oral proficiency. *ELT Journal* 40(3):212-220.

Skehan, P. (1989). *Individual Differences in Second-Language Learning*. London: Edward Arnold.

Stansfield, C. & Hansen, J. (1983). Field dependence-independence as a variable in second language cloze test performance. *TESOL Quarterly* 17(1):29-38.

van Lier, L. (1989). Reeling, writhing, fainting and stretch in coils: oral proficiency interviews as conversation. *TESOL Quarterly* 23(3):489-508.

Zammit, S.A. (1993). Motivation, test results, gender differences and foreign languages: how do they connect? Paper presented at the 15th Language Testing Research Colloquium, Arnhem and Cambridge. August 1993.

Zeidner, M. (1986). Are English language aptitude tests biased towards culturally different minority groups? Some Israeli findings. *Language Testing* 3(1):80-95.

Zeidner, M. (1987). A comparison of ethnic, sex and age biases in the predictive validity of English language aptitude tests: some Israeli data. *Language Testing* 4(1):55-71.